

ME 731 - Métodos em Análise Multivariada
Segundo semestre de 2015
Lista de Exercícios IV

1. Resolva todos os exercícios deixados em sala.
2. Considere $\mathbf{X} \sim D_2(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\rho \in (-1, 1)$. Responda os itens:
 - a) Calcule os autovalores e autovetores (ortonormalizados) associados à $\mathbf{\Sigma}$.
 - b) Obtenha as componentes principais associadas ao vetor \mathbf{X} .
 - c) Interprete as componentes principais que você encontrou no item b).
 - d) O que ocorre com as componentes principais (em termos de suas variâncias e dos autovetores associados à elas) quando $|\rho| \rightarrow 1$. Como as variáveis originais passam a estar relacionadas às componentes principais.
 - e) Repita os itens a), b), c), d) quando $\rho = 0$.
3. Considere $\mathbf{X} \sim D_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, $\mathbf{\Sigma}$ uma matriz positiva definida, com autovalores - autovetores (ortonormalizados) $(\lambda_1, \mathbf{e}'_1)', (\lambda_2, \mathbf{e}'_2)', \dots, (\lambda_p, \mathbf{e}'_p)'$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Prove que a variância generalizada de $\mathbf{\Sigma}$ é igual ao produto dos autovalores de $\mathbf{\Sigma}$. Discuta as vantagens e desvantagens de se utilizar a variância generalizada ao invés da variância total (soma das variâncias) como um critério de escolha do número de componentes principais.
4. Com base no conjunto de dados sobre as tartarugas (Questão 4 da Lista III) e considerando a matriz de correlações, utilize a metodologia das componentes principais. Responda os seguintes itens:
 - a) Calcule os autovalores e autovetores (ortonormalizados) associados à matriz de correlações.
 - b) Faça o “screeplot” e construa uma tabela com a variância explicada por cada componente e acumulada. Quantas componentes você reteria? Justifique, adequadamente, sua resposta.
 - c) Calcule as correlações entre cada uma das componentes e as variáveis originais, interpretando os resultados.
 - d) Interprete cada uma das componentes que você obteve. A suposição de normalidade é razoável para alguma(s) dela(s)?

- e) Considere as duas primeiras componentes obtidas. A variância explicada por elas é razoável? Justifique sua resposta. Faça um bi-plot, em que os “labels” são os gêneros, interpretando os resultados.
 - f) Com apenas a primeira componente, compare os dois grupos, em termos de suas médias, utilizando uma metodologia que você ache adequada, escrevendo as conclusões pertinentes.
5. Com base no conjunto de dados sobre informações nutricionais de cereais (o arquivo está disponível na página do curso sob o nome de *cereal.txt*), e considerando a matriz de correlações, utilize a metodologia das componentes principais para responder os seguintes itens (quando pertinente):
- a) Faça uma análise descritiva dos dados originais, por fabricante.
 - b) Calcule os autovalores e autovetores (ortonormalizados) associados à matriz de correlações.
 - c) Faça o “screeplot” e construa uma tabela com a variância explicada por cada componente e acumulada. Quantas componentes você reteria? Justifique adequadamente sua resposta.
 - d) Calcule as correlações entre cada uma das componentes e as variáveis originais, interpretando os resultados.
 - e) Interprete cada uma das componentes que você obteve. A suposição de normalidade é razoável para alguma(s) dela(s)?
 - f) Considere as duas primeiras componentes obtidas. A variância explicada por elas é razoável? Justifique sua resposta.
 - g) Com as duas primeiras componentes faça uma análise descritiva, por fabricante, interpretando os resultados. Faça um bi-plot, em que os “labels” são os fabricantes, interpretando os resultados.
 - h) Com apenas a primeira componente, compare os fabricantes, em termos de suas médias, utilizando uma metodologia que você ache adequada, escrevendo as conclusões pertinentes.
6. Dados relativos à moscas chamadas de “bitting fly” (banco de dados moscas). Os dados se referem à sete variáveis medidas em duas espécies desse tipo de mosca, a saber *Leptoconops carteri* e *Leptoconops torrens* (de agora em diante, *carteri* e *torrens*, por simplicidade). As variáveis (na ordem em que aparecem no banco de dados) são: espécie (0 - *torrens* e 1- *carteri*), comprimento da asa, largura da asa, comprimento do terceiro palpo, largura do terceiro palpo, comprimento do quarto palpo, comprimento do 12º

segmento da antena e comprimento do 13^o segmento da antena. Realize uma análise de componentes principais e, com base nas três primeiras componentes (justificando adequadamente a escolha) faça uma análise dos dados, levando em consideração os grupos, conforme visto em sala de aula (não se esqueça de fazer o biplot). Além disso, utilizando a primeira componente principal, compare os grupos, através de um modelo de regressão adequado (fazendo uma análise residual para o modelo ajustado). Compare as conclusões que você obteve nesta questão com aquelas obtidas na Questão 1).